

How to make the most out of the trunk and branch wgMLST schema for *Brucella* spp.

The power of a trunk and branch schema

This schema provides you with a powerful tool to study *Brucella* spp. both on genus and species level. The advantage of it being a “trunk and branch” schema is that it shows the full diversity of the relative homogenous *Brucella* while obtaining a high and uniform genome coverage for all species.

The trunk is a fixed set of loci that share the same CDS structure across the genus. This is not a core schema since these trunk loci do not need to be conserved. Around 60% of all loci in this schema are assigned to the trunk. Additionally, for each species, loci were added to obtain a subschema (branch) that provides an optimal set of loci for your species-specific analysis. To conserve maximal consistency with classical typing methods, separate subschemas are available with respectively 9⁽¹⁾ and 21 MLST loci⁽²⁾. The result is a genus-wide schema with 14 subschemas (see Fig. 1).

What are the species-specific loci?

Besides having some unique genes, every species in the genus *Brucella* has a set of CDSs that differ in structure - e.g. later start, different frame, different stop, etc. Both the unique and the structurally different CDSs were determined for every species, resulting in species-specific subschemas that consist of the trunk and a set of loci (see Fig. 1). This does not mean that individual loci cannot be shared amongst different subschemas. The structure of a CDS might only be different for 1 or 2 species and identical for the other ones.

Does this change anything to the way you use the schema?

The submission of your samples to the Calculation Engine and the way BioNumerics retrieves the results remain exactly the same. The imported results are the allele calls on all loci present in the schema. To make optimal use of this genus-wide schema however, the subschema you use for the analysis workflow will depend on the final results you want to obtain.

In practice: two possible situations with examples

1. You want to compare multiple species

1a. You want to see where your sample fits in between the diversity of reference genomes. Use the “trunk” subschema.

1b. You have samples from multiple sources/species and want to know the relation between them. Use the “trunk” subschema.

2. You are interested in one species

2a. You know which species you isolated. Use the species-specific subschema.

2b. You don't know which species you isolated.

First, use the “trunk” subschema to compare your sample to a series of reference strains (either your own historical data or our demo database). Then, use the corresponding species-specific subschema to compare them to closely related samples (e.g. potential outbreak samples).

Tutorials and movies can be found on our website by clicking [this](#) direct link.

Schema 1

Trunk (3246 loci)

Schema 2 and 3

9 loci MLST

21 loci MLST

Schema 4 to 14: species-specific (= trunk + loci listed below)

+850 loci *B. abortus*

+937 loci *B. canis*

+873 loci *B. ceti*

+876 loci *B. inopinata*

+971 loci *B. melitensis*

+856 loci *B. microti*

+822 loci *B. neotomae*

+864 loci *B. ovis*

+887 loci *B. pinnipedialis*

+945 loci *B. suis*

+645 loci *B. vulpis*

Fig. 1. Structure of the BioNumerics *Brucella* spp. wgMLST schema

References:

(1) Whatmore, A. M., Perrett, L. L., & MacMillan, A. P. (2007). Characterisation of the genetic diversity of *Brucella* by multilocus sequencing. BMC microbiology, 7(1), 34.

(2) Whatmore, A. M., Koylass, M. S., Muchowski, J., Edwards-Smallbone, J., Gopal, K. K., & Perrett, L. L. (2016). Extended Multilocus Sequence Analysis to Describe the Global Population Structure of the Genus *Brucella*: Phylogeography and Relationship to Biovars. Frontiers in Microbiology, 7.